

# Preconditioning ideas for the Augmented Lagrangian method

A. M. Sajo-Castelli<sup>a</sup>

<sup>a</sup>*Departamento de Cómputo Científico y Estadística, Universidad Simón Bolívar, Ap. 89000, Caracas 1080-A, Venezuela.*

---

## Abstract

A preconditioning strategy for the Powell-Hestenes-Rockafellar Augmented Lagrangian method (ALM) is presented. The scheme exploits the structure of the Augmented Lagrangian Hessian. It is a modular preconditioner consisting of two blocks. The first one is associated with the Lagrangian of the objective while the second administers the Jacobian of the constraints and possible low-rank corrections to the Hessian. The proposed updating strategies take advantage of ALM convergence results and avoid frequent refreshing. Constraint administration takes into account complementarity over the Lagrange multipliers and admits relaxation. The preconditioner is designed for problems where constraint quantity is small compared to the search space. A virtue of the scheme is that it is agnostic to the preconditioning technique used for the Hessian of the Lagrangian function. The strategy described can be used for linear and nonlinear preconditioning. Numerical experiments report on spectral properties of preconditioned matrices from Matrix Market while some optimization problems were taken from the CUTEst collection. Preliminary results indicate that the proposed scheme could be attractive and further experimentation is encouraged.

*Keywords:* Augmented Lagrangian Method, Preconditioning, Iterative methods.

*2000 MSC:* 65F10, 15A12, 65F35, 90C30, 90C25.

---

## 1. Introduction

Augmented Lagrangian methods (ALM) are practical and affordable algorithms extensively used in applied fields. They are designed to solve large-scale nonlinear optimization problems possibly with nonlinear constraints. The constraints are classified in two groups: hard and soft. Hard constraints are such that strict fulfillment is required in order to accept a solution while minor infeasibility for the soft constraints is granted. Birgin and Martínez in [1] present a recent and detailed overview of practical ALM. These methods are considered a general optimization machinery [2, 3, 4, 5, 6, 1, 7] in the sense that they successfully cope with a great variety of real-life problems. ALM has also been adapted to very specific applications [8, 9, 10]. A great virtue of the method is that it can be accelerated via preconditioning techniques.

Iterative Krylov-type methods have been used inside ALM and even though they are very well studied and can handle very large-scale problems, it is well known the poor convergence speed. Accelerating these methods by preconditioning strategies is common practice. In the last 50 years a considerable amount of effort has been invested in the design and construction of effective preconditioners. In the context of ALM, there has not been much motivation to study acceleration strategies that benefit or exploit convergence results, although recently some studies on very special problems present non-induced preconditioning that benefit from ALM convergence [11, 12, 13].

State of the art ALM implementations that can be highlighted are Algencan [1, 3, 4] and Lancelot B [14, 15]. Both are considered production-grade codes that use a variety of direct and iterative methods, such as the Conjugate Gradients method. In regards to acceleration, Algencan does not offer enough flexibility while Lancelot B incorporates a list facilities, it can also leave to the user the task of administering the whole preconditioning process. Surely this last option covers all possible scenarios, but it can also be daunting for non-expert users. From the user-land perspective, these MLA implementations leave a certain void preconditioning-wise. The proposed acceleration scheme tries to fill this gap.

In this work we present a modular acceleration scheme that exploits the special structure of the Augmented Lagrangian of Powell-Hestenes-Rockafellar. A key aspect is that the update strategies take advantage of ALM

---

*Email address:* [asajo@usb.ve](mailto:asajo@usb.ve) (A. M. Sajo-Castelli)

convergence. The scheme has two main ingredients, an *auxiliary* preconditioner associated with the Lagrangian function and a machinery that administers the constraints. Separating these components has the great advantages of freedom in choosing the auxiliary preconditioner and having full control over the Jacobian matrix of the constraints and possible low-rank corrections. Such corrections are attractive because they promote quality in the approximation to the Hessian of the Lagrangian function. The proposed scheme is considered a generalization of the QNCGNA preconditioner used in Algencan [16, 1].

The rest of this document is organized as follows. We briefly introduce the problem of interest and the Augmented Lagrangian Method followed by the presentation of the acceleration scheme. The preconditioner is introduced in two parts. We start by showing how to accelerate ALM for problems with a single constraint, then a general preconditioner is presented. Update strategies for the various components are discussed and some illustrative numerical results are reported. The work concludes with some final remarks.

## 2. The Augmented Lagrangian Method

Let us consider the following optimization problem

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && h(x) = 0, \\ & && g(x) \leq 0, \\ & && x \in \Omega, \end{aligned} \tag{1}$$

where  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  are continuous functions, in particular  $f(\cdot)$  is two times continuously differentiable and  $\Omega = \{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$ . With the goal to solve (1), two functions are associated, the Lagrangian

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^m h_i(x) \lambda_i + \sum_{i=1}^p g_i(x) \mu_i = f(x) + h(x)^\top \lambda + g(x)^\top \mu, \tag{2}$$

where  $\lambda \in \mathbb{R}^m$  y  $\mu \in \mathbb{R}_+^p$  are the *Lagrange multipliers* of the problem, and the *Augmented Lagrangian* of Powell-Hestenes-Rockafellar [17, 18]

$$L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \sum_{i=1}^m \left[ h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \frac{\rho}{2} \sum_{i=1}^p \left[ \max \left( 0, g_i(x) + \frac{\mu_i}{\rho} \right) \right]^2, \tag{3}$$

with the *external penalty parameter*  $\rho > 0$ ,  $\lambda \in \mathbb{R}^m$  y  $\mu \in \mathbb{R}_+^p$ . Taking the Augmented Lagrangian function as the objective, then solving (1) and minimizing (3) with respect to  $x \in \Omega$ , are equivalent problems. The ALM consists of solving a sequence of *sub-problems* and updating the the Lagrange multipliers and the external penalty parameter as required. This naturally divides the iterations in two groups, external or Lagrangian iterations and internal iterations. On the external iterations the multipliers and the penalty parameters are updated while the internal iterations are dedicated in solving the sub-problem. Soft constraints are *moved up* to the objective penalizing a shifted version of infeasibility measure. Hard constraints are enforce inside the inner solver which is dedicated to the sub-problem

$$\begin{aligned} & \text{minimize} && L_{\rho_k}(x, \lambda_k, \mu_k) \\ & \text{subject to} && x \in \Omega, \end{aligned} \tag{4}$$

where  $\rho_k$ ,  $\lambda_k$  and  $\mu_k$  are fixed.

The ALM is conceptually presented in Algorithm 1. The algorithm is schematic and leaves open the choice on how to solve the sub-problem (4), it only requires that  $x_k$  be an *approximate* solution. On each external iteration (Step 4) it is considered if  $x_k$  has done enough progress in regards to feasibility and complementarity. In cases where progress is good, the external penalty parameter does not need to change, on the contrary the parameter must be incremented. The idea behind the shifts  $\lambda/\rho$  and  $\mu/\rho$  is that even when the external penalty parameter has a *moderate* penalty, there can exist *adequate* values for the shifts on the multipliers where it is possible to find an acceptable solution to (4). Given the interest of the algorithm to produce a numerically attractive result, on Step 5 the multipliers are safeguarded.

**Input:**  $\lambda_{\min} < \lambda_{\max}, \mu_{\max} > 0, \gamma > 1, 0 < \tau < 1, \bar{\lambda}_1 \in [\lambda_{\min}, \lambda_{\max}]^m, \bar{\mu}_1 \in [0, \mu_{\max}]^p, \rho_1 > 0. k \leftarrow 1.$   
**Step 1** Find approximate solution  $x_k \in \mathbb{R}^n$  of (4).  
**Step 2** Stop if  $x_k$  satisfies (6)–(8).  
**Step 3** Update Lagrange multipliers:

$$\lambda_{k+1} = \bar{\lambda}_k + \rho_k h(x_k) \quad \text{and} \quad \mu_{k+1} = \left( \bar{\mu}_k + \rho_k g(x_k) \right)_+$$

**Step 4** Update  $\rho$ :

Let  $V_{i,k} = \min \left( -g_i(x_k), \bar{\mu}_{i,k}/\rho_k \right), \quad i = 1 : p.$

**if**  $k = 1 \vee \max \left\{ \|h(x_k)\|, \|V_k\| \right\} \leq \tau \max \left\{ \|h(x_{k-1})\|, \|V_{k-1}\| \right\}$  **then**

$\rho_{k+1} \geq \rho_k.$

**else**

$\rho_{k+1} = \gamma \rho_k.$

**end if**

**Step 5** Safeguard multipliers:  $\bar{\lambda}_{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m, \bar{\mu}_{k+1} \in [0, \mu_{\max}]^p.$

**Step 6** Increment iteration counter  $k \leftarrow k + 1$  and go to Step 1.

Algorithm 1: Augmented Lagrange Method [1].

We briefly present convergence results for Algorithm 1, details and theorem proofs are given in [1, Chapter 5]. For the following results it is assumed that on Step 1 of Algorithm 1,  $x_k$  is a global minimizer of (4):

$$L_{\rho^k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq L_{\rho^k}(x, \bar{\lambda}^k, \bar{\mu}^k) + \varepsilon^k, \quad \forall x \in \Omega,$$

where  $\{\varepsilon^k\} \subseteq \mathbb{R}_+$  is bounded and  $\varepsilon^k$  need not be *small*.

**Theorem 1 (Feasibility).** *Let  $\{x^k\}$  be a sequence generated by Algorithm 1 under the previous assumption and let  $x^*$  be a limit point of this sequence. Then,*

$$\|h(x^*)\|_2^2 + \|g(x^*)_+\|_2^2 \leq \|h(x)\|_2^2 + \|g(x)_+\|_2^2, \quad \forall x \in \Omega.$$

This result guarantees that if the problem is feasible, then every limit point of the sequence generated by ALM is also feasible.

**Theorem 2 (Optimality).** *Let  $\{x^k\}$  be a sequence generated by Algorithm 1 and  $x^*$  be a limit point of  $\{x^k\}$ . Suppose that  $\varepsilon^k \xrightarrow{k \rightarrow \infty} 0$  and that the problem (4) is feasible. On Step 4, when possible,  $\rho_{k+1}$  is not incremented. Then,  $x^*$  is a global minimizer.*

For practical purposes it is not necessary to differentiate between the equality and inequality constraints, they can be both considered under the *umbrella* constraint function  $c(x)$  and the associated multiplier vector  $\lambda$ . Optimality conditions and constraint handling requires only minor attention. The Augmented Lagrangian becomes

$$F(x) = L_\rho(x, \lambda) = f(x) + \frac{\rho}{2} \sum_{i \in \mathbb{E}} \left[ c_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \frac{\rho}{2} \sum_{i \in \mathbb{I}} \left[ \max \left( 0, c_i(x) + \frac{\lambda_i}{\rho} \right) \right]^2 \quad (5)$$

where the sets  $\mathbb{E}$  and  $\mathbb{I}$  contain the indexes of equality and inequality constraints respectively.

Supposing that problem (1) has solution, then the pair  $(x_k, \lambda_k)$  is minimizer of (5) —and in consequence is also of (3) and solution to (1)— when it satisfies

$$\left\| P_\Omega \left( x_k - \left[ \nabla L_{\rho_k}(x_k, \lambda_k) + \sum_{j=1}^m \lambda_k \nabla c_j(x_k) \right] \right) - x_k \right\| \leq \varepsilon_{\text{opt}}, \quad (6)$$

$$\max \left( \max_{j \in \mathbb{E}} \left( |c_j(x_k)| \right), \max_{j \in \mathbb{I}} \left( |\min(-c_j(x_k), \lambda_k)| \right) \right) \leq \varepsilon_{\text{fact}}, \quad (7)$$

$$\max \left( \max_{j \in \mathbb{E}} \left( |c_j(x_k)| \right), \max_{j \in \mathbb{I}} \left( c_j(x_k)_+ \right) \right) \leq \varepsilon_{\text{fact}}, \quad (8)$$

where

$$\lambda_k = \begin{cases} \lambda_{j,k-1} + \rho_k c_j(x_k) & \text{if } j \in \mathbb{E} \text{ or } \lambda_{j,k-1} + \rho_k c_j(x_k) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The operator  $P_\Omega$  is the Euclidean projection on the convex set  $\Omega$ . Conditions (6) and (7) insure that  $(x_k, \lambda_k)$  is a stationary point of the problem while condition (8) guarantees that the point satisfies the required feasibility tolerance.

It is advantageous to differentiate between Lagrangian and internal iterations. External iterates and counter are identified by  $x_k$  and  $k$  respectively while internal ones use  $z_\ell$  and  $\ell$  respectively. If the set  $\Omega = \mathbb{R}^n$  then the sub-problem is said to be *unconstrained*, on the contrary it will be assumed to be *convex constrained*.

Practical Newton-type methods are very popular for solving unconstrained problems and require a descent direction of first order given by solving the quadratic model

$$H_\ell d_\ell = -\nabla F(z_\ell),$$

where  $H_\ell$  is the Hessian  $\nabla^2 F(z_\ell)$  or an approximation to it. For the Truncated-Newton method,  $d_\ell$  is estimated using the Conjugate Gradients method in the SPD case or Minimal Residual method when  $H_\ell$  is symmetric but undefined. These methods are quite attractive if accelerated with high quality preconditioners.

Practical solvers for convex constrained problems are the Spectral Projected Gradient method (SPG) [19, 20, 21] and its preconditioned variant (PSPG) [22]. PSPG has received some attention [23] and can be viewed as a nonlinear preconditioned variant that combines the Preconditioned Spectral Gradient method and SPG. For these type of problems we assume that  $P_\Omega$  is the Euclidean projection operator over the convex set  $\Omega$ , exists and is of acceptable cost. We also require that the first order derivatives of  $f(\cdot)$  and  $c(\cdot)$  exist wherever required.

### 3. Preconditioning ideas

Under our context, accelerating ALM may refer to accelerate the resolution of the sub-problem (Newton-type directions) as well as the acceleration for the estimation of the descent direction (Cauchy-type machinery). Given the nature of the method to solve at each external iteration an optimization problem, and that the sequence of these problems tend to be similar, preconditioning schemes must *recycle* between iterations in order to be attractive.

In order to propose preconditioning schemes for (4), it is necessary to study the explicit form of the Hessian of the objective (Augmented Lagrangian function). For historical reasons, most applications use and implement the Lagrangian function and not its augmented counterpart. Although, as we shall see both are closely related. The Lagrangian function, gradient and Hessian associated to (5), are

$$\begin{aligned} \mathcal{L}(z, \lambda) &= f(z) + \sum_{i=1}^m c_i(z) \lambda_i = f(z) + c(z)^\top \lambda, \\ \nabla \mathcal{L}(z, \lambda) &= \nabla f(z) + \sum_{i=1}^m \lambda_i \nabla c_i(z) = \nabla f(z) + \nabla c(z) \lambda, \quad \text{with} \quad \nabla c(z) = \begin{bmatrix} \left| \nabla c_1(z) \right| & \left| \nabla c_2(z) \right| & \cdots & \left| \nabla c_m(z) \right| \end{bmatrix}, \\ \nabla^2 \mathcal{L}(z, \lambda) &= \nabla^2 f(z) + \sum_{i=1}^m \lambda_i \nabla^2 c_i(z). \end{aligned}$$

On the other hand, the Augmented Lagrangian counterpart is

$$\begin{aligned} L_\rho(z, \lambda) &= f(z) + \frac{\rho}{2} \sum_{i=1}^m \left[ c_i(z) + \frac{\lambda_i}{\rho} \right]^2 = f(z) + c(z)^\top \lambda + \frac{\rho}{2} c(z)^\top c(z) + \frac{1}{2\rho} \lambda^\top \lambda, \\ \nabla L_\rho(z, \lambda) &= \nabla f(z) + \sum_{i=1}^m [\lambda_i + \rho c_i(z)] \nabla c_i(z) = \nabla f(z) + \nabla c(z) [\lambda + \rho c(z)], \\ \nabla^2 L_\rho(z, \lambda) &= \nabla^2 f(z) + \sum_{i=1}^m \lambda_i \nabla^2 c_i(z) + \rho \sum_{i=1}^m c_i(z) \nabla^2 c_i(z) + \rho \sum_{i=1}^m \nabla c_i(z) \nabla c_i(z)^\top, \end{aligned}$$

grouping  $\nabla^2 c_i(z)$  and doing the change:  $\hat{\lambda} = \lambda + \rho c(z)$ , we have

$$\nabla L_\rho(z, \lambda) = \nabla^2 f(z) + \sum_{i=1}^m \hat{\lambda}_i \nabla^2 c_i(z) + \rho \nabla c(z) \nabla c(z)^\top.$$

Noting that  $\nabla^2 \mathcal{L}(z, \hat{\lambda}) = \nabla^2 f(z) + \sum_{i=1}^m \hat{\lambda}_i \nabla^2 c_i(z)$ , we obtain the key identity

$$\nabla^2 L_\rho(z, \lambda) = \nabla^2 \mathcal{L}(z, \hat{\lambda}) + \rho \nabla c(z) \nabla c(z)^\top, \quad \text{or equivalently,} \quad \nabla^2 F(z_\ell) = H_\ell = M_\ell + \rho V_\ell V_\ell^\top.$$

This matrix sum has many relevant characteristics. For instance, since  $M$  is the sum of Hessian matrices it is symmetric and close to a solution is definite [24]. Under certain choices of  $M$ , it can be verified that it is always definite [6]. The Gauss-Newton matrix  $VV^\top$  is always symmetric and can be regarded semi-definite. Moreover, if  $m < n$  then the rank of  $VV^\top$  is at best  $m$ , the number of soft constraints. In practice, methods exploit the complementary condition in such a way that the rank of  $VV^\top$  is at most the number of *active non-relaxed* constraint count at current iterate.

When solving problem (4) using Newton-type directions, it is necessary to solve the quadratic model

$$\nabla^2 F(z) d = H d = [M + \rho V V^\top] d = -\nabla F(z). \quad (9)$$

For Cauchy-type directions, if  $P$  is a preconditioner for  $H$ , then we interpret the preconditioner  $P$  as an approximation to  $H^{-1}$  and the *enriched* descent direction is

$$d = -P \nabla F(z).$$

Contrasting the previous two expressions, applying nonlinear preconditioning for Gradient-type methods fortunately is analogous as accelerating the Newton-type machinery. Our interest lies in solving the linear system (9) using Krylov-type iterative methods, PCG when  $H$  is PD or MinRes for the undefined case. Preconditioning schemes for ALM must be able to at least exploit the following two desirable key features:

1. *Preconditioner recycling.* The idea is to take advantage of convergence for the sub-problems. Supposing that (1) has solution, then we expect  $H_k \xrightarrow{k \rightarrow \infty} H^*$ , meaning that starting from a certain iterate  $k$  (or  $\ell$ ) it is possible to bound the difference  $H_{k+1} - H_k$  and successfully use the same preconditioner onwards.
2. *Preconditioner update.* Assembly of  $P_{k+1}$  should partially reuse work invested in assembling  $P_k$ . Generally speaking this is not straightforward. As an illustrative example on the involved difficulties, low-rank updates [25, 16, 26, 27, 28] are highlighted with special focus on BFGS-type corrections [29, 30].

In what follows we propose a new preconditioning scheme that takes advantage of these features.

### 3.1. Preconditioning singly constrained sub-problems

Many applications can be modeled using singly constrained problems: support vector machine formulations for classification or pattern recognition [31, 32] are but two mainstream examples. This subsection introduces a new inverse preconditioner scheme for solving the following singly constrained problem

$$\begin{aligned} \text{minimize} \quad & F(z) = L_\rho(z, \lambda_1) = f(z) + \frac{\rho}{2} \left[ c_1(z) + \frac{\lambda_1}{\rho} \right]^2 \\ \text{subject to} \quad & z \in \Omega. \end{aligned} \quad (10)$$

We simplify notation by relaxing the index from the constraint and its associated Lagrange multiplier. With the idea to introduce the preconditioner, suppose a Newton-type machinery is used to solve (10), then at each internal iteration  $\ell$  it is required to find  $d$  from the linear system

$$H_\ell d = -\nabla F(z_\ell),$$

with

$$H_\ell = \nabla^2 L_{\rho_k}(z_\ell, \lambda_k) = \nabla^2 \mathcal{L}(z_\ell, \lambda_k + \rho_k c(z_\ell)) + \rho_k \nabla c(z_\ell) \nabla c(z_\ell)^\top = M_\ell + \rho_k v_\ell v_\ell^\top.$$

Relaxing the iteration indexes, we have

$$[M + \rho vv^\top] d = -\nabla F(z), \quad (11)$$

where  $v = \nabla c(z)$  is considered a non-null vector,  $vv^\top$  is a rank-1 matrix and we suppose  $M$  is full rank. Given our interest in solving (11) using PCG or MinRes, applying the preconditioner  $P^{-1}$  [24, Algorithm 5.3: line 3 and equation (5.38d)] reduces to efficiently compute the product

$$h_j = P^{-1}r_j, \quad \text{with} \quad r_0 = H z_0 + \nabla F(z_0).$$

The matrix in (11) is special in the sense that it is the sum of an invertible matrix plus a rank-1 matrix. Suppose that  $P^{-1}$  is precisely  $H^{-1} = [M + \rho vv^\top]^{-1}$ , then the product  $h_j = P^{-1}r_j$  can be computed using the Sherman-Morrison identity [33],

$$\begin{aligned} h_j &= [M + \rho vv^\top]^{-1} r_j, \\ &= \left[ M^{-1} - \frac{1}{1 + \rho v^\top M^{-1} v} \rho M^{-1} v v^\top M^{-1} \right] r_j = M^{-1} r_j - \frac{1}{1 + \rho v^\top M^{-1} v} \rho M^{-1} v v^\top M^{-1} r_j, \end{aligned}$$

solving for  $a$ ,  $Ma = r_j$  or  $a = M^{-1}r_j$ , we obtain

$$h_j = a - \frac{\rho}{1 + \rho v^\top M^{-1} v} M^{-1} v v^\top a,$$

analogously we find  $b$ ,  $Mb = v$  or  $b = M^{-1}v$  and observing that  $v^\top a$  is a scalar, we finally have

$$h_j = a - \frac{\rho v^\top a}{1 + \rho v^\top b} b.$$

If  $M$  is not trivially invertible, finding  $a$  and  $b$  requires an *auxiliary* preconditioner  $P_M$ . Each form of obtaining  $a$  and consequently  $b$ , give rise the different preconditioning strategies. It is noteworthy to remark that the preconditioner  $P$  is never assembled and is considered an *abstract* preconditioner which relies upon the auxiliary preconditioner  $P_M$ . This highlights the *agnostic* nature of  $P$ , for it does not enforce any specific choice over  $P_M$ . Under this scheme, the spectrum of the preconditioned matrix  $P^{-1}[M + \rho vv^\top]$  is described by the following result.

**Theorem 3.** *Suppose  $M$  is the Hessian matrix of the Lagrangian function associated to (10), let the preconditioner for  $H$  be  $P = [M + \rho vv^\top]$ , furthermore let  $P_M^{-1} > 0$  be a preconditioner for  $M$ ,  $\rho \geq 1$  and  $\mathcal{E}_M := P_M^{-1}M - I$ , then the spectrum of  $P^{-1}H$  is*

$$\Lambda(P^{-1}H) = \Lambda\left(I + (1 - \rho) P_M^{-1} v v^\top \mathcal{E}_M\right), \quad \text{where} \quad v = \frac{\rho}{1 + \rho v^\top P_M^{-1} v}.$$

The proof of this theorem can be found in [34]. This result has a few important implications. If  $\rho \rightarrow \infty$ , then  $P^{-1}H \rightarrow I$ . This indicates that for very large values of  $\rho$  it is not attractive or even necessary to precondition. It also shows that the condition of  $P^{-1}H$  can be described in terms of  $P_M^{-1}M$ , this is to say that the quality of  $P$  is given in direct relation to the quality of  $P_M$ .

### 3.2. General Preconditioner

In this section we work with the problem

$$\begin{aligned} \text{minimize} \quad & F(z) = L_\rho(z, \lambda) = f(z) + \frac{\rho}{2} \sum_{i=1}^m \left[ c_i(z) + \frac{\lambda_i}{\rho} \right]^2 \\ \text{subject to} \quad & z \in \Omega, \end{aligned} \quad (12)$$

where the quantity of constraints is less than the problem dimension. Analogous to (11), the linear system to solve is

$$Hd = [M + \rho VV^\top] d = -\nabla F(z),$$

where  $V = \nabla c(z)$  is the constraint Jacobian,  $VV^\top$  is the Gauss-Newton matrix of rank at most  $m$  and we suppose that  $M$  is rank complete. Initially, we set the preconditioner to  $P^{-1} = [M + \rho VV^\top]^{-1}$ . The use of the Sherman-Morrison-Woodbury [35, 36] inverse closed formula rises many difficulties, specially the stringent requirement of a composed matrix inverse which for our case is not practical. Fortunately,  $H$  can be rewritten as a recursion over the columns of  $V$ , suppose (12) has three constraints ( $m = 3$ ), then

$$H = \left( ((M + \rho v_1 v_1^\top) + \rho v_2 v_2^\top) + \rho v_3 v_3^\top \right).$$

This formulation shows that  $H$  can always be written as an invertible matrix plus a rank-1 matrix. This is also true for the preconditioner  $P$ , let  $P_0 := M$ , then

$$P_1 = P_0 + \rho v_1 v_1^\top, \quad P_2 = P_1 + \rho v_2 v_2^\top, \quad \text{and finally,} \quad P = P_3 = P_2 + \rho v_3 v_3^\top,$$

which leads to the following recursion,

$$P = P_m = P_{m-1} + \rho v_m v_m^\top, \text{ with } P_0 := M, \text{ and } m \geq 1.$$

Using Miller's inverse formula [37], the product  $h_j = P^{-1}r_j$  is reduced to the recursion

$$h_j = P^{-1}r_j = P_3^{-1}r_j = [P_2 + v_3 v_3^\top]^{-1}r_j, \quad (13)$$

but this was previously shown how to be solved, and is

$$h_j = a_4 = a_3 - \frac{\rho v_3^\top a_3}{1 + \rho v_3^\top b_3} b_3, \text{ with } a_3 = P_2^{-1}r_j \text{ and } b_3 = P_2^{-1}v_3.$$

Now, in a similar fashion we compute  $a_i$ ,  $i = 3 : 1$ :

$$\begin{aligned} a_3 &= P_2^{-1}r_j = a_2 - \frac{\rho v_2^\top a_2}{1 + \rho v_2^\top b_2} b_2, \text{ with } a_2 = P_1^{-1}r_j \text{ and } b_2 = P_1^{-1}v_2, \\ a_2 &= P_1^{-1}r_j = a_1 - \frac{\rho v_1^\top a_1}{1 + \rho v_1^\top b_1} b_1, \text{ with } a_1 = P_0^{-1}r_j \text{ and } b_1 = P_0^{-1}v_1, \\ a_1 &= M^{-1}r_j. \end{aligned}$$

Unfortunately, the estimation of the  $b_i$ ,  $i = 3 : 1$  is a bit more elaborate: in order to compute  $b_3$  it is required to have  $P_2^{-1}v_3$ , which in turn requires  $P_1^{-1}v_3$  and  $P_1^{-1}v_2$ ,

$$\begin{aligned} b_3 &= P_2^{-1}v_3 = a'_2 - \frac{\rho v_2^\top a'_2}{1 + \rho v_2^\top b_2} b_2, \text{ with } a'_2 = P_1^{-1}v_3 \text{ and } b_2 = P_1^{-1}v_2, \\ b_2 &= P_1^{-1}v_2 = a'_1 - \frac{\rho v_1^\top a'_1}{1 + \rho v_1^\top b_1} b_1, \text{ with } a'_1 = P_0^{-1}v_2 \text{ and } b_1 = P_0^{-1}v_1, \\ b_1 &= M^{-1}v_1. \end{aligned}$$

Implementation-wise, the  $a'_i$  are computed at the same time as the  $b_i$ ,  $i = 2 : 1$  which leads to a secondary recursion.

Although the previous formulation requires double recursion over the columns of  $V$ , it will be shown to have many attractive features. The elements  $a_i, b_i, a'_i$  and finally the acceleration product  $h_j$ , are computed by applying multiple times the Sherman-Morrison identity. This is considered a very attractive aspect because it only requires Matrix-Vector and internal products.

In the same spirit as the single constrained case, each form of obtaining  $a_i$  and  $b_i$  give rise to different preconditioners and this variant is also considered to be an agnostic acceleration scheme which leaves open the strategy on how to estimate  $M^{-1}a_i$  and  $M^{-1}b_i$ . This has numerous advantages, we highlight the fact that handling the Lagrangian independently from the constraint Jacobian permits to exploit the sparse structure of the problem.

In the next section we generalize the previous example of three constraints.

### 3.2.1. The $B$ Matrix

Observing the recursion that computes the product (13), it can be seen that elements  $a'_i$  and  $b_i$  *do not* depend on  $r_j$  but on the columns of  $V$ . This in a natural way induces to *pre-compute*  $a'_i$  and  $b_i$ , and save them efficiently in a *B storage matrix*. In what follows we show how to efficiently compute the product  $h = P^{-1}r$ .

Let  $h_0 := M^{-1}r$ ,  $B_{0,*} = M^{-1}V$  where  $B_{i,j}$  indicates the  $j$ -th column of matrix  $B_i$ .

On a first pass, storage matrix  $B$  is assembled,

$$B_{i,j} = B_{i-1,j} - \frac{\rho v_i^\top B_{i-1,j}}{1 + \rho v_i^\top B_{i-1,i}} B_{i-1,i}, \quad \text{for } j = i : m, \quad \text{with } i = 1 : m.$$

On a second pass,  $h$  is computed,

$$h_i = h_{i-1} - \frac{\rho v_i^\top h_{i-1}}{1 + \rho v_i^\top B_{i-1,i}} B_{i-1,i}, \quad \text{for } i = 1 : m.$$

Finally  $h = h_m$ . It is important to note that from the implementation point of view, the elements of  $B_{i+1,*}$  overwrite those of  $B_{i,*}$  in such a way as not to waste storage. The recursion to compute  $h$  distills to

$$h_i = h_{i-1} - \frac{\rho v_i^\top h_{i-1}}{1 + \rho v_i^\top B_i} B_i,$$

where  $B_i$  regains its classical meaning indicating the  $i$ -th column of  $B$ . Storage matrix  $B$  unites in a single matrix all the ingredients to estimate  $h$  and as such it is intimately related to  $M$  and  $V$ . The previous formulation efficiently computes the product  $P^{-1}w$  or  $[M + \rho VV^\top]^{-1}w$  for any  $w$  as long as  $M$  and  $V$  stay relatively the same. Significant changes in  $V$  and/or  $M$  force re-assembly of  $B$ . This fact outlines certain updating aspects that an acceleration scheme must be aware of. It is evident that this strategy to compute  $h$  is only attractive when  $m$  is far from  $n$ . The proposed preconditioner  $P$  can be shown trivially that is adequate for use within PCG and MinRes.

### 3.2.2. Secant type directions

In large-scale applications computing the Augmented Lagrangian Hessian is a luxury seldom available, in general terms a reasonable approximation is used. Within our context, the Augmented Lagrangian Hessian has clear differentiation between its components

$$\nabla^2 L_\rho(z, \lambda) = \nabla^2 f(z) + \sum_{i=1}^m [\lambda_i + \rho c_i(z)] \nabla^2 c_i(z) + \rho \sum_{i=1}^m \nabla c_i(z) \nabla c_i(z)^\top = \nabla^2 f(z) + C(z) + V(z).$$

Krejić et al. [6] propose using  $\nabla^2 f(z) + V(z)$  and Birgin and Martínez suggest in [1] to use only  $V(z)$  along with two *corrections*. These involve a spectral correction [38, 39, 19, 40] using the associated Rayleigh quotient [41, 42] and a second correction [16, 1] in the spirit of BFGS that forces to satisfy the Secant equation [27, 26]

$$H_\ell s_\ell = y_\ell,$$

$$\hat{H}_\ell = \nabla^2 f(z_\ell) + V(z_\ell), \quad s = z_\ell - z_{\ell-1}, \quad y_\ell = \nabla F(z_\ell) - \nabla F(z_{\ell-1}).$$

These corrections are low-rank [28, 43, 44, 30, 45, 46] and are crafted with the main purpose of having a closed inverse form. For illustrative purposes,  $\hat{H}$  is corrected spectrally and with the famous BFGS formula,

$$\begin{aligned} \hat{H}_+ &:= \hat{H} + \sigma I, \quad \text{with } \sigma = \frac{(y - \hat{H}(z_\ell) s)^\top s}{s^\top s}, \\ H &= \hat{H}_+ + \frac{yy^\top}{s^\top y} - \frac{\hat{H}_+ s s^\top \hat{H}_+}{s^\top \hat{H}_+ s}, \quad \text{if } s^\top y \neq 0. \end{aligned}$$

Now, if  $P^{-1} \approx H^{-1}$ , the assembly and use of the preconditioner is analogous as shown at the start of §3.2 since  $H$  can be rewritten as an invertible matrix plus matrices of rank-1. Let us see,

$$\begin{aligned} H &= \hat{H}_+ + \frac{yy^\top}{s^\top y} - \frac{\hat{H}_+ s s^\top \hat{H}_+}{s^\top \hat{H}_+ s}, \quad \text{noting that } \hat{H}_+ \text{ is symmetric, we have} \\ &= \nabla^2 f(z) + V(z_\ell) + \sigma I + \nu yy^\top - \psi ww^\top, \quad \nu := \frac{1}{s^\top y}, \quad w := \hat{H}_+ s, \quad \psi := \frac{1}{s^\top w} \\ &= \nabla^2 f(z) + \sigma I + VV^\top + \nu yy^\top - \psi ww^\top, \end{aligned}$$



letting  $M = \nabla^2 f(z) + \sigma I$ , we have

$$H = M + VV^\top + \nu yy^\top - \psi ww^\top.$$

That is,  $H$  has the form required to assemble  $P^{-1}$ . The matrix  $M$  is guaranteed to be rank complete by means of the first spectral correction. With some abuse of notation, the constraint Jacobian is *augmented* to accommodate the elements associated with the BFGS correction

$$V \leftarrow \begin{bmatrix} | & | & | \\ \sqrt{\rho} V & \sqrt{\nu} y & \sqrt{\psi} w \\ | & | & | \end{bmatrix}, \quad \text{signs} := [1, 1, \dots, 1, 1, -1]^\top.$$

The auxiliary vector “signs” is used inside the recursion associated with  $B$ .

### 3.2.3. Special Exact Case

When using Quasi-Newton directions, the choice of the approximation to the Lagrangian Hessian plays a key role in determining the quality of the auxiliary preconditioner associated to  $M$ . If the election of  $M$  that approximates  $\nabla^2 \mathcal{L}(z)$  has explicit inverse, then  $P^{-1}$  is the exact (theoretical) inverse of  $H$ . A trivial case is to choose  $M$  diagonal. As a side-effect under this context, the proposed preconditioner can be seen as a generalization of the QNCGNA preconditioner of Birgin and Martínez [16, 1]. The preconditioner is designed to work on the linear system

$$V(z_\ell) d = -g_\ell.$$

The matrix  $V(z)$  is symmetric and SPD, it is corrected spectrally and, if possible, a second BFGS-style correction is applied. The corrected approximation to the Augmented Lagrangian Hessian is

$$H = \begin{cases} V_+ := V(z) + \sigma I & \text{if } y^\top s < 10^{-8} \|y\| \|s\|, \\ V_+ + \frac{yy^\top}{s^\top y} - \frac{V_+ s s^\top V_+}{s^\top V_+ s} & \text{on the contrary,} \end{cases} \quad (14)$$

and the QNCGNA preconditioner

$$P_{\text{QNCGNA}} = \begin{cases} D_+ := \text{diag}(V(z)) + \sigma_D I & \text{if } y^\top s < 10^{-8} \|y\| \|s\|, \\ D_+ + \frac{yy^\top}{s^\top y} - \frac{D_+ s s^\top D_+}{s^\top D_+ s} & \text{on the contrary.} \end{cases}$$

Witch has explicit closed inverse

$$P_{\text{QNCGNA}}^{-1} = D_+^{-1} + \frac{(s - D_+^{-1}y)s^\top + s(s - D_+^{-1}y)^\top}{s^\top y} - \frac{(s - D_+^{-1}y)^\top y s s^\top}{(s^\top y)^2}.$$

Note that the search direction is done over  $V(z)$  and the preconditioner is over  $\text{diag}(V(z))$ .

Now taking up (14), the matrix  $H$  can be rewritten as a rank complete matrix plus the sum of rank-1 matrices. Suppose the BFGS correction is possible, then

$$H = V(z) + \sigma I + \frac{yy^\top}{s^\top y} - \frac{V_+ s s^\top V_+}{s^\top V_+ s}, \quad V_+ = V(z) + \sigma I.$$

Noting that  $V_+$  is symmetric, we have

$$\begin{aligned} &= V(z) + \sigma I + \frac{yy^\top}{s^\top y} - \frac{V_+ s (V_+ s)^\top}{s^\top V_+ s}, \\ &= \sigma I + \rho VV^\top + \nu yy^\top - \psi ww^\top, \quad \nu = \frac{1}{s^\top y}, \quad w = V_+ s, \quad \psi = \frac{1}{s^\top w}. \end{aligned}$$

Let  $M = \sigma I$  and  $V$  —again, with some notation abuse— be the matrix of size  $n \times (m+q)$  that gathers the constraint Jacobian and the  $q$  vectors associated with the second BFGS correction, then defining trivially  $P_M = \sigma^{-1} I$  we have that  $P^{-1}$  is the explicit closed inverse of  $H$ .

### 3.3. Update Strategies

There exist a variety of updating choices for the preconditioner within ALM. Given the available granularity inside the recursive nature of applying the preconditioner ( $P^{-1}r_k$ ), update strategies enjoy a fine-grain control over each component. Practical update strategies monitor changes on  $\|M_\ell - M_{\ell-1}\|$  and  $\|V_\ell - V_{\ell-1}\|$  independently and take the following actions:

1. Update  $P_M$  whenever  $\|M_\ell - M_{\ell-1}\|_1 > \delta_M$ .
2. Update  $B$  each time  $\|V_\ell - V_{\ell-1}\|_1 > \delta_V$  or  $\|M_\ell - M_{\ell-1}\|_1 > \delta_M$ .
3. Apply *relaxation* over the columns of  $V$  when assembling  $B$  or at the moment of applying the preconditioner ( $P^{-1}r_k$ ). This can be done by using only those columns of  $V$  that in norm are greater than a threshold or that have huge infeasibility measure,  $V = (v_i)$ ,  $i = 1 : m$  |  $\|v_i\| > \varepsilon_v > 0 \vee i \in \mathbb{E}$  |  $|c_i(x)| > \varepsilon_c > 0 \vee i \in \mathbb{I}$  |  $(c_i(x))_+ > \varepsilon_c$ .

The idea behind the first item is that if  $M_\ell$  and  $M_{\ell+1}$  are similar, then possibly  $P_{M_\ell}$  will also be a good preconditioner for  $M_{\ell+1}$ . Item two establishes that  $B$  should only be updated if  $V_{\ell+1}$  and  $V_\ell$  greatly differ. It also forces an update if  $P_{M_\ell}$  was updated. Up until not finding the final search space,  $V$  will be changing drastically and update schemes must be aware of this. The last item is based on the idea that small elements should have small contributions and can be safely discarded.

#### 3.3.1. Strategies for $P_M$

The proposed scheme leaves open the choice on how to precondition  $M$ . Clearly the updating of  $P_M$  depends on such choice. Nevertheless some maintenance aspects of  $P_M$  can be mentioned. The update of  $P_M$  should be delayed as much as possible due to the fact that updates on  $P_M$  force an update on  $B$ . A great advantage on the modularity of  $P$  is that it allows among other things, to change preconditioning strategy ( $P_M$ ) mid-way between two updates.

Preliminary experimentation shows that most of the big changes for  $\|M_\ell - M_{\ell-1}\|_1$  occur at the beginning and specially between two external iterations due to the Lagrangian multipliers and the external penalty parameters being updated. Given the convergence of ALM and supposing the problem has solution, iterates will converge asymptotically to a point where updating  $P_M$  will no longer be necessary. The reported experiments suggest to use a lax threshold for the update of  $P_M$ . This choice promotes frequent updates only at the beginning of the resolution while avoiding unnecessary updates towards the end.

#### 3.3.2. On the update of matrix $B$

Matrix  $B$  is associated with the constraint Jacobian and as such its form is described by the active non-relaxed constraints and the rank-1 correction artifacts. Given the recursive nature of the assembly of  $B$ , it is possible to establish predictive update strategies that reduce costs. The idea is the following. Let  $V_{m-2}$  be the second to last column of  $V$  and the gradient associated with the inequality constraint  $c_{m-2}(z_\ell)$  and also suppose that it has a very small infeasibility measure. Most probably the next iterate will inactivate this constraint forcing the discarding of  $V_{m-2}$ . This induces an update *only* to the last two columns of  $B$ . Now, if  $V_{m-2}$  were to be the first column of  $V$  then the induced update would affect *all* the columns of  $B$  having a much greater cost. This observation suggests orderings over the columns of  $V$  that potentially reduce the cost of updating  $B$ . A practical ordering could be induced by the infeasibility measure of each constraint

$$|c_i(z)|, \quad i \in \mathbb{E}, \quad \text{y} \quad \max(0, c_i(z)), \quad i \in \mathbb{I}.$$

The spirit behind this ordering is to leave for the end of the recursion those columns whose associated constraints will (possibly) soon be discarded. This ordering could be further improved by taking into account the norm of each associated gradient. It is also very convenient to leave at the end of the recursion the  $q$  columns associated with BFGS-type corrections. These elements can have a vivid transit state since initially iterations may or may not fulfill the condition  $s^T y < 10^{-8} \|s\| \|y\|$ . Although, it has been observed that close to a solution the BFGS correction can always be applied. Numerical experimentation shows that as iterates approach a solution, changes in  $V$  diminish down to a point where updating  $B$  is not necessary and can be recycled successfully.

### 3.4. Comments

Accelerating Cauchy-type methods require to consider preconditioning matrices as approximations to the inverse of the Hessian. Acceleration enriches the descent direction  $-\nabla F(x)$  with second order information from the approximation  $H$  of  $\nabla^2 F(x)$ . In others words, the enriched direction is obtained by solving

$$H_k d_k = -\nabla F(x_k),$$

but this is exactly the same task as applying acceleration under Newton-like choices. Hence the instructions on how to apply and when to update the (linear) preconditioner are analogous for the nonlinear case. Unfortunately the enriched (preconditioned) direction is not always a descent direction. In these cases, the enriched direction is discarded in favor of  $-\nabla F(x_k)$ , albeit, the work invested in building the approximation  $P \approx H^{-1}$  should not be discarded: it could potentially be recycled the next time an enriched direction is to be computed.

The assembly and maintenance of matrix  $B$  loses appeal and stops being attractive in the presence of a large number of active constraints, say  $m \approx n$  or even  $m > n$ . In this scenario, we have  $H = M + W$  where  $W$  is dense and possibly rank complete or near complete. As iterates start closing-in to a solution, the number of active constraints should decrease down to a point where the use of  $B$  is practical. This suggest to handle acceleration for problems with a large amount of constraints in a two-stage approach.

A possible heuristic is to use Quasi-Newton directions induced by  $V_{\mathbb{K}}(z)$  where  $\mathbb{K}$  is the set of constraints indexes for the  $0 < K \ll n$  elements with greatest infeasibility measure. The quadratic model for the direction is

$$V_K V_K^T d = -\nabla F(z), \quad \text{or even} \quad [M + V_K V_K^T] d = -\nabla F(z).$$

Under these two choices, it is possible to use the proposed scheme. It is important to note that before finding the final search space, the set  $\mathbb{K}$  will be changing inducing unfavorable frequent updates on matrix  $B$ .

It is important to consider alternative techniques that do not use the explicit form of the constraint Jacobian but can tackle with the dense Gauss-Newton matrix  $W$ . This topic is considered open for future study.

## 4. Numerical Experiments

All experiments were run using Matlab® R2012a on an Intel® Core™ i7-2640M CPU @ 2.80GHz with 8 GB of memory. We start by examining the quality of  $P$  as a preconditioner for  $H$  and its efficiency at solving linear systems of the form  $Hx = y$ . Some experiments involving constraint relaxation and preconditioner update strategies follow. We finalize by solving unconstrained and box-constrained problems from the CUTEst[47] data-set.

### 4.1. Spectral Properties of the Preconditioned Matrix

We wish to understand the spectral properties of the preconditioned matrix  $P^{-1}H = P^{-1}[M + \rho VV^T]$ . Quality metric is based on condition number and spectra of  $P^{-1}H$ . For these experiments, the auxiliary preconditioner  $P_M$  is of the family of Robust Incomplete Factorization of type SAINV [48, 49, 50, 51] and is considered a black-box that executes the matrix-vector product  $M^{-1}Y$ . Table 1 and Figure 1 report two particular experiments on sparse random matrices of size 100 with  $m \in \{10, 50\}$  random constraints. These experiments use a diverse range of values for the external penalty  $\rho$  and dropping  $\tau$  parameters. From the table and figure it can be observed that the higher the quality of  $P_M$  (smaller  $\tau$ ) the spectrum of  $P^{-1}H$  accumulates around the identity, albeit, if  $P_M$  is poor then the conditions of  $P^{-1}H$  and  $H$  are comparable. Curiously for these problems, higher values of  $\rho$  seem to have a favorable effect over the condition of  $P^{-1}H$ . Increased values of  $\rho$  worsen the condition of  $H$  while promoting the one of  $P^{-1}H$ .

The next set of experiments involve solving the linear system of equations  $Hx = y$  with  $H = M + \rho VV^T$  using the Conjugate Gradients method. Tested matrices are sparse random and from the Matrix Market [52] collection. All matrices are forced to be real SPD. Constraints are random with  $N(0, 1)$  distribution. Convergence tolerance is set to  $10^{-8}$ . Obtained result are reported in Table 2. Dropping parameters  $\tau_1$  y  $\tau_2$  are associated with  $P_M$ . Column labeled with “ $m$ ” represents the number of constraints and the columns associated with  $\text{nzz}(Z)$  give an idea of the density of  $P_M$ . Columns “CG” and “PCG” report the number of iterations required. From the table an evident influence of  $\rho$  over the conditioning of  $H$  is observed. The higher the quality of  $P_M$ , the more evident is the acceleration. On some problems, for very large values  $\rho$  the number of CG iterations is surprisingly low and preconditioning stops being practical. Conversely, in some cases where  $H$  is very ill conditioned, preconditioning not only is very effective but is the only alternative that converges. In PCG context, increment in  $\rho$  generally implies reduction in iteration count.

Table 1: Experimental results for solving the linear system  $Hx = y$ , with  $H = M + \rho VV^T$  for random  $M \in \mathbb{R}^{100 \times 100}$  and  $V \in \mathbb{R}^{100 \times m}$ .

$m$	$\rho$	$\tau$	$\kappa_1(H)$	$\kappa_1(P^{-1}H)$
10	1.5	0.1	$2.4 \times 10^5$	$5.4 \times 10^4$
10	15.5	0.1	$5.8 \times 10^5$	$4.9 \times 10^4$
10	154.8	0.1	$4.1 \times 10^6$	$4.9 \times 10^4$
10	1548.3	0.1	$4.0 \times 10^7$	$4.9 \times 10^4$
10	15483	0.1	$4.0 \times 10^8$	$4.9 \times 10^4$
10	1.5	0.001	$2.4 \times 10^5$	1.3
10	15.5	0.001	$5.8 \times 10^5$	1.3
10	154.8	0.001	$4.1 \times 10^6$	1.3
10	1548.3	0.001	$4.0 \times 10^7$	1.3
10	15483	0.001	$4.0 \times 10^8$	1.3

$m$	$\rho$	$\tau$	$\kappa_1(H)$	$\kappa_1(P^{-1}H)$
50	1.5	0.1	$2.0 \times 10^4$	$5.5 \times 10^3$
50	15.5	0.1	$7.3 \times 10^4$	$3.4 \times 10^3$
50	154.8	0.1	$6.5 \times 10^5$	$3.6 \times 10^3$
50	1548.3	0.1	$6.4 \times 10^6$	$3.7 \times 10^3$
50	15483	0.1	$6.4 \times 10^7$	$3.7 \times 10^3$
50	1.5	0.01	$2.0 \times 10^4$	20
50	15.5	0.01	$7.3 \times 10^4$	13
50	154.8	0.01	$6.5 \times 10^5$	12
50	1548.3	0.01	$6.4 \times 10^6$	12
50	15483	0.01	$6.4 \times 10^7$	12

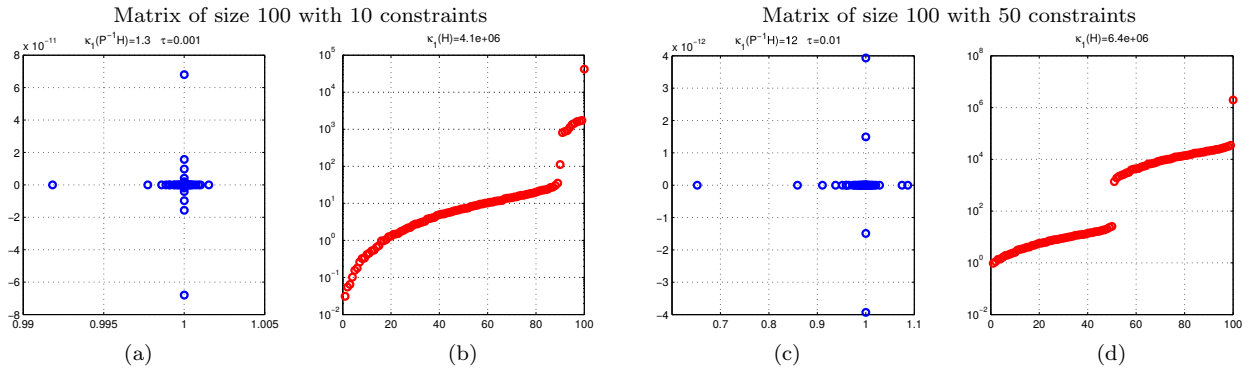


Figure 1: Spectral distribution of  $P^{-1}H$  and  $H$ . Figures (a) and (c), show in blue color the eigenvalues of  $P^{-1}H$  grouped near the identity  $(1, 0)$ . Figures (b) and (d) show in red color the eigenvalue distribution of  $H$  in logarithmic scale.

Table 2: Experimental results for solving the linear system  $Hx = y$  using Conjugate Gradients method.

Name	$n$	$m$	$\tau_1$	$\tau_2$	$\text{nzz}(Z)$	$\frac{\text{nzz}(Z)}{n^2}$	$\frac{\text{nzz}(Z)}{\text{nzz}(M)}$	$\rho$	$\kappa(H)$	$\kappa(P^{-1}H)$	CG	PCG
Sparse	1000	1	0.05	0.075	2040	0.002	0.674	1	12681509	74245	n/c	22
Sparse	1000	1	0.05	0.075	2040	0.002	0.674	100	1259615534	74251	172	14
Sparse	1000	1	0.05	0.075	2040	0.002	0.674	1000	12595389913	74251	58	9
Sparse	1000	100	0.1	0.1	1656	0.002	0.684	100	153329943	12	56	2
Sparse	1000	50	0.1	0.1	1666	0.002	0.684	1	2650446	122	173	9
Sparse	1000	50	0.1	0.1	1666	0.002	0.684	100	264040355	121	54	3
Sparse	2000	1	0.05	0.075	2865	0.001	0.746	1	70457448	70219	n/c	15
Sparse	2000	1	0.05	0.075	2865	0.001	0.746	1000	70285577870	70214	37	3
bcsprw01	49	20	0.1	0.1	131	0.086	1	1	208	1.3	21	5
bcsprw01	49	20	0.1	0.1	131	0.086	1	10000	1841063	1.2	27	2
bcsprw01	49	1	0.18	0.18	142	0.059	0.85	1	171	15	31	17
bcsprw01	49	1	0.18	0.18	142	0.059	0.85	100	13474	15	30	14
bcsprw01	49	1	0.18	0.18	142	0.059	0.85	10000	1343884	15	21	9
bcsprw01	49	1	0.18	0.18	142	0.059	0.85	100000	13438526	15	9	6
bcsprw02	49	1	0.18	0.18	159	0.066	0.952	1	600	53	33	18
bcsprw02	49	1	0.18	0.18	159	0.066	0.952	100	45618	51	32	15
bcsprw02	49	1	0.18	0.18	159	0.066	0.952	10000	4547969	51	20	12
bcsprw02	49	1	0.18	0.18	159	0.066	0.952	1000000	454783100	51	5	3
bcsprw02	49	10	0.1	0.1	157	0.065	0.94	1	294	1.6	25	6
bcsprw02	49	10	0.1	0.1	157	0.065	0.94	10000	2253720	1.4	27	2
bcsprw02	49	30	0.1	0.1	157	0.065	0.94	1	355	1.3	21	5
bcsprw02	49	30	0.1	0.1	157	0.065	0.94	10	2720	1.2	39	4
bcsstm03	112	1	0.7	0.7	112	0.009	1	1	3003.3	1	42	1
bcsstm03	112	1	0.7	0.7	112	0.009	1	1000000	23186836	1	19	1
bcsstm03	112	1	0.7	0.7	112	0.009	1	10000000	231867364	1	7	1

#### 4.2. Constraint Relaxation & Update

Updating the preconditioner on each iteration is prohibitively expensive and even unnecessary. The convergence of ALM guarantees that sub-problems will tend to be similar up to the point where recycling the preconditioner is possible. For inequality constraints, as soon as the iterate is feasible or strongly feasible, the number of active constraints drops drastically. This leads to propose cheap update strategies that discard elements from the  $B$  matrix. Active constraint relaxation must be done with certain care. Preliminary experiments suggest it is not easy to establish the *contribution* of each active constraint to the Gauss-Newton matrix. With the idea of understanding the influence of each constraint (column of  $V$ ) on the quality of the preconditioner  $P$ , spectral properties of  $P^{-1}H$  are monitored while assembling  $P$  with a sub-set of columns of  $V$ . For these experiments  $P_M$  can be considered  $M^{-1}$ . Preconditioner  $P$  is assembled by increasing the amount of columns of  $V$  included. The order of inclusion is given by the norm of each column,  $\|v_1\| \geq \|v_2\| \geq \dots \geq \|v_m\| > 0$ . Figure 2 shows the condition number of  $P^{-1}H$  in function of  $\rho$  and the number of columns of  $V$  used in the assembly of  $P$ . Poor behavior can be observed, in order to maintain good quality most of the components of  $V$  must be used. Attractive relaxation strategies have to account at least for the norm and infeasibility measure of the components of  $V$ . This suggests to assemble  $P$  using only those active constraints where  $\|\sqrt{\rho_k} \nabla c_i(x_\ell)\| \geq \varepsilon_v$  and  $|c_i(x_\ell)| \geq \varepsilon_c$ .

Updating strategies for  $P$  must take advantage of the convergence of ALM. At first changes to  $M$  and  $V$  are assumed to be *big* but will tend to smooth as the sub-problems start to be similar. In order to understand the associated costs of diverse updating strategies for  $P$ , the following experiments are conducted. Let  $\delta_M$  and  $\delta_v$  be updating tolerances for changes in matrices  $M$  and  $V$  between two consecutive iterations. We solve problem C4 [34] varying parameters  $\delta_M$ ,  $\delta_v$ ,  $\varepsilon_v$  (relaxation on  $\nabla c(x)$ ) and  $\tau$  (dropping tolerance for  $P_M$ ), while observing the amount of updates required to produce a solution. Table 3 reports a sequence of experiments on problem C4 to understand updating strategies/cost relation. The cost of each configuration is given by CG/MinRes iterations and the amount of updates on  $M$  and  $V$  required to find a solution. These two counters are antagonistic. In general terms, low CG/MinRes iteration count corresponds to a high update count for  $M$  and  $V$ . The idea is to find a compromise between these two. Reported numerical results in Table 3 give an intuitive overview of attractive configurations.

Table 4 shows in detail the dynamics of certain parameters while solving problem C4 using QN machinery.

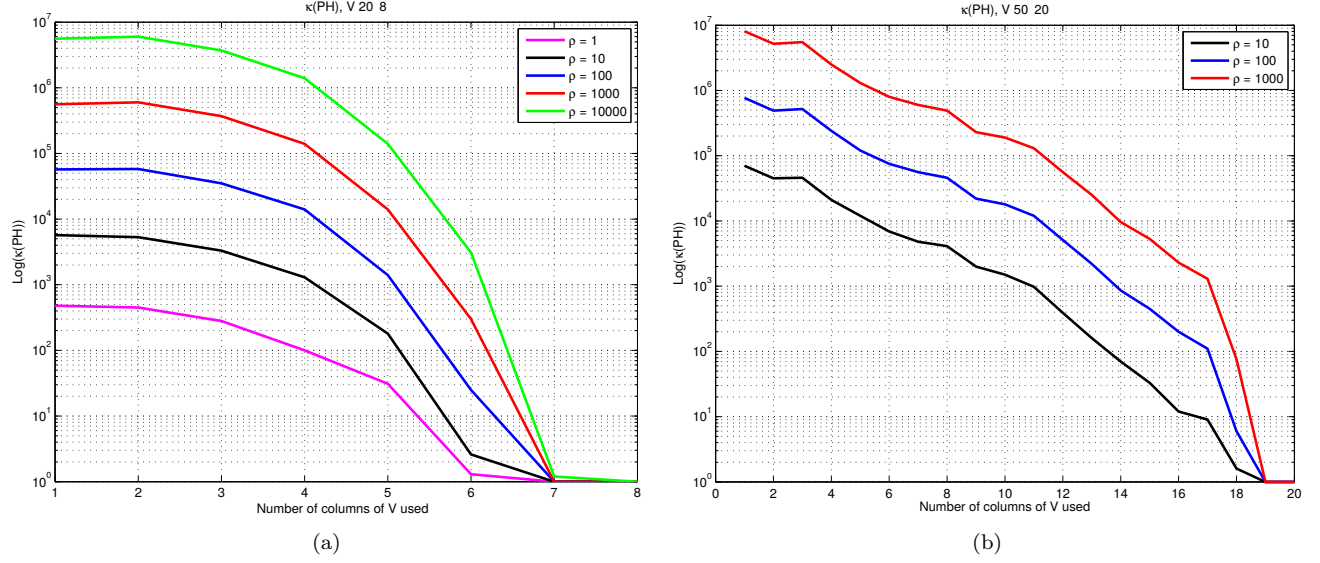


Figure 2: Accumulative influence of the columns of  $V$  in the quality of the preconditioner  $P$ . In (a) the problem is of size 20 with 8 active constraints. For  $P$  to be competitive, the first 9 columns of  $V$  are required in the assembly of  $B$ . In (b) the problem has dimension 50 with 20 constraints. In this case the first 19 columns of  $V$  are needed.

Table 3: Numerical experiments to understand iteration costs based on update strategies and modulation of the parameters  $\delta_M$ ,  $\delta_v$ ,  $\varepsilon_v$  and  $\tau$  for problem C4.

Strategy	$\delta_M$	$\delta_v$	$\varepsilon_v$	$\tau$	Ext. it	Int. it	GC/MR it	Updates		
								Total	$M$	$V$
Auto	$1 \cdot 10^{-1}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$10^{-6}$	37	98	317	56	20	36
Auto	$1 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$10^{-6}$	37	98	317	62	20	42
Auto	$1 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-3}$	$10^{-6}$	37	98	317	62	20	42
Auto	$1 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-3}$	$10^{-9}$	37	98	317	62	20	42
Auto	$1 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-3}$	$10^{-6}$	37	98	258	65	30	35
Auto	$2 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-3}$	$10^{-9}$	37	98	345	62	18	44
Auto	$2 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$5 \cdot 10^{-3}$	$10^{-9}$	37	98	345	62	18	44
Auto	$5 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$5 \cdot 10^{-3}$	$10^{-6}$	37	98	368	62	17	45
Auto	$5 \cdot 10^{-1}$	$5 \cdot 10^{-1}$	$1 \cdot 10^{-1}$	$10^{-3}$	37	98	399	27	17	10
Auto	$5 \cdot 10^{-1}$	$5 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$10^{-6}$	37	98	394	27	17	10
Auto	$5 \cdot 10^{-1}$	$5 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$10^{-3}$	37	98	373	65	17	48
Auto	$5 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-3}$	$10^{-9}$	37	98	284	62	22	40
Update every external iteration					37	98	533	35	3	32
Assemble $P$ only once					37	98	1156	1	1	0
Newton / direct method					37	98	—	—	—	—

The problem is rigged to have ten variables with nine inequality and one equality constraints. Descent direction is found using MinRes. The auxiliary preconditioner  $P_M$  is ILU of type Crout with dropping tolerance  $10^{-7}$ . Column labeled “Update type” specifies the three reasons for updating  $P$ : “M” and “V” indicate that changes between two consecutive  $M$  or  $V$  matrices where greater that  $\delta_M$  and/or  $\delta_v$  respectively, while “V\*” indicates a *forced* update to  $H$  due to BFGS-style corrections ( $|\mathbb{V}|$  changes from 10 to 12). It can also indicate that the correction cannot be applied in the current iteration ( $|\mathbb{V}|$  changes from 12 to 10).  $|\mathbb{V}|$  indicates the quantity of columns of matrix  $V$  in use and  $\text{rnnz}(Y)$  indicates the density of  $P_M$  relative to  $M$ . From Table 4 it can be observed that initially the changes between two consecutive matrices  $M$  and  $V$  are substantial, hence the update type “MV”. Further in the resolution of the problem, changes to  $M$  start to attenuate then diminish for  $V$ . In general terms, this behavior can be considered characteristic. In particular, for problem C4 after iteration 4 of Lagrangian iteration 3 it was no longer necessary to update  $P$  which represents a 44.8% reduction in unnecessary updates. Obtained results show

Table 4: Evolution of the iterates while solving problem C4 using ALM using QN-type directions.

L. it	N. it	PMR it (MR it)	Update type	$ \mathbb{V} $	$\rho$	$\text{rnnz}(Y)$	$\ M_k - M_{k-1}\ _1$	$\ v_k - v_{k-1}\ _1$
1	1	1 (10)	MV	10	100	1.6	—	—
1	2	1 (7)	MV*	12	100	1.6	$1 \times 10^3$	$1.16 \times 10^3$
1	3	1 (5)	MV	12	100	1.6	13.9	788
1	4	1 (5)	MV	12	100	1.6	2.95	850
1	5	1 (5)	MV	12	100	1.6	$9.42 \times 10^{-1}$	318
1	6	1 (5)	MV	12	100	1.6	$5.37 \times 10^{-1}$	354
1	7	1 (5)	MV	12	100	1.6	$7.13 \times 10^{-1}$	129
1	8	1 (5)	MV	12	100	1.6	$1.56 \times 10^{-1}$	189
1	9	1 (7)	MV	12	100	1.6	978	50.3
1	10	1 (8)	MV	12	100	1.6	9.81	2.1
2	1	1 (10)	MV*	10	100	1.6	12.6	0.0
2	2	1 (10)	MV*	12	100	1.6	2.59	4.91
3	1	1 (10)	MV*	10	100	1.6	2.53	0.0
3	2	1 (10)	M	10	100	1.6	$1.38 \times 10^{-1}$	0.0
3	3	3 (10)	V*	12	100	1.6	$3.56 \times 10^{-3}$	3.76
3	4	3 (10)	V	12	100	1.6	$2.22 \times 10^{-3}$	$3.77 \times 10^{-1}$
3	5	3 (10)	—	12	100	1.6	$1.11 \times 10^{-3}$	$9.78 \times 10^{-3}$
3	6	3 (10)	—	12	100	1.6	$5.53 \times 10^{-4}$	$6.36 \times 10^{-3}$
3	7	3 (10)	—	12	100	1.6	$2.76 \times 10^{-4}$	$3.20 \times 10^{-3}$
3	8	3 (10)	—	12	100	1.6	$1.38 \times 10^{-4}$	$1.60 \times 10^{-3}$
3	9	3 (10)	—	12	100	1.6	$6.90 \times 10^{-5}$	$8.02 \times 10^{-4}$
3	10	3 (10)	—	12	100	1.6	$3.45 \times 10^{-5}$	$4.01 \times 10^{-4}$
3	11	3 (10)	—	12	100	1.6	$1.72 \times 10^{-5}$	$2.01 \times 10^{-4}$
3	12	3 (10)	—	12	100	1.6	$8.62 \times 10^{-6}$	$1.00 \times 10^{-4}$
3	13	3 (10)	—	12	100	1.6	$4.31 \times 10^{-6}$	$5.02 \times 10^{-5}$
3	14	2 (10)	—	12	100	1.6	$2.16 \times 10^{-6}$	$2.51 \times 10^{-5}$
3	15	2 (10)	—	12	100	1.6	$1.08 \times 10^{-6}$	$1.25 \times 10^{-5}$
3	16	2 (10)	—	12	100	1.6	$5.39 \times 10^{-7}$	$6.27 \times 10^{-6}$
3	17	2 (10)	—	12	100	1.6	$5.39 \times 10^{-7}$	$2.89 \times 10^{-6}$

the convenience of updating  $P$  via  $\delta_M$  and  $\delta_v$  tolerances over strategies that update on predefined iterations e. g. every Lagrangian iteration. Unfortunately, experimentation showed that attractive update tolerances is problem dependent and requires individual adjustment. For some problems these can be lax and still give good acceleration results, specially for changes between elements of  $V$ . It is also attractive to relax the *small* columns of  $V$  since it potentially reduces costs and can efficiently be controlled using a single additional relaxation parameter. Adding and removing columns from  $V$  is a frequent task. This requires cheap machineries that handle efficiently the addition and removal of elements within the matrix  $B$ . Furthermore, the assembly order inside of  $B$  can take advantage of

a prediction ingredient based on constraint gradient size and infeasibility measure.

#### 4.3. Unconstrained Sub-problems

In each Lagrangian iteration, the problem to solve is

$$\text{minimize } F(x) = L_{\rho_k}(x, \lambda_k) \text{ subject to } x \in \mathbb{R}^n, \text{ with } \rho_k \text{ and } \lambda_k \text{ fixed.}$$

Two type of quadratic models where used for the direction. Newton (NW)  $H_\ell = \nabla^2 F(z_\ell)$  and Quasi-Newton (QN)

$$H_\ell = \nabla^2 f(x_\ell) + \sum_{i \in \mathbb{A}} \nabla c_i(x_\ell) \nabla c_i(x_\ell)^T + \sigma I + \text{BFGS},$$

where  $\mathbb{A}$  is the set of active non-relaxed constraints indexes,  $\sigma$  the inverse Rayleigh quotient and BFGS represents rank-one corrections. The auxiliary preconditioner  $P_M$  used is from the ILU family. Table 5 reports the obtained experimental results. Acceleration is evident when contrasting columns labeled “Itpd” and “ItL”. For this set of problems, the preconditioning scheme achieved acceleration factors between 1.3 and 9.6. On some problems

Table 5: Numerical results for solving unconstrained problems using ALM coupled with Newton-type methods. Columns labeled by “ItL”, “Itin”, “Itpd” and “ItL” indicate the number of external, internal, MinRes and accelerated MinRes iteration count. The sum of columns “AcM” and “AcV” reports the times the preconditioner  $P$  was updated while each column shows individual contribution.

Name	$n$	$m$	Method	ItL	Itin	Itpd	ItL	AcM	AcV	Time (s)
C4-1	10	10	NW	11	52	82	420	35	1	2.4
C4-1	10	10	NW	10	53	83	432	34	2	2.5
C4-1	10	10	QN	45	143	311	1365	40	19	5.7
C4-1	10	10	QN	45	143	176	1365	107	6	5.9
C4-R-10	10	10	QN	3	29	55	252	14	2	1.0
C4-R-20	20	10	QN	3	48	265	617	33	6	1.6
C4-R-25	25	10	QN	4	55	112	900	37	18	1.7
<hr/>										
BT3	5	3	NW	8	84	84	807	1	3	1.3
BT3	5	3	NW	7	86	86	539	1	1	1.4
BT3	5	3	QN	8	129	182	651	21	58	1.8
BT3	5	3	QN	8	129	193	651	20	50	1.7
<hr/>										
BT8	5	2	NW	5	72	87	148	27	0	1.1
BT8	5	2	QN	5	88	108	169	26	8	1.3
<hr/>										
BT11	5	3	NW	10	140	322	707	38	4	2.1
BT11	5	3	NW	8	114	134	680	112	0	1.8

the QN direction did not converge. Most of the computational effort is invested in updating  $P_M$ , this promotes approximations to  $M$  that are easy invertible, such as band-diagonal approximations. Preconditioners used in the NW and QN models are also used in the Spectral Gradient method (SG)<sup>1</sup> and its preconditioned variant (PSG). Comparing the column “Itin” of the rows “SG” and “PSG” for each problem in Table 6 acceleration in iteration count is evident, but if comparing computational cost, preconditioning is not always attractive.

#### 4.4. Box-constrained Sub-problems

For these experiments, the problem to solve is

$$\text{min } F(x) = L_{\rho_k}(x, \lambda_k) \text{ subject to } \ell_i \leq x \leq u_i, \text{ with } \rho_k \text{ and } \lambda_k \text{ fixed.}$$

In order to understand the efficiency of the scheme, the sub-problems are solved using SPG and its preconditioned variant PSPG. Table 7 reports numerical experiments, similar results to Table 6 were obtained. Acceleration is clear

<sup>1</sup>SPG code adapted from the TANGO project [19, 20].



Table 6: Numerical results for solving unconstrained problems using ALM coupled with SG and it preconditioned variant PSG. Column labels are described in the caption of Table 5.

Name	$n$	$m$	Method	ItL.	Itin	AcM	AcV	Time (s)
C4-1	10	10	SG	n/c	—	—	—	—
C4-1	10	10	PSG	1	21	3	0	0.3
BT3	5	3	SG	9	486	—	—	1.6
BT3	5	3	PSG	7	19	12	0	0.4
BT8	5	2	SG	3	29	—	—	0.3
BT8	5	2	PSG	3	17	5	0	0.3
BT9	4	2	SG	9	193	—	—	0.8
BT9	4	2	PSG	9	32	25	0	0.5
BT11	5	3	SG	7	426	—	—	1.5
BT11	5	3	PSG	7	28	28	0	0.4
BT12	5	3	SG	7	1222	—	—	4.3
BT12	5	3	PSG	6	16	12	0	0.3
HS48	5	2	SG	2	131	—	—	0.5
HS48	5	2	PSG	1	2	2	0	0.2
MAKELA4	21	40	SG	2	5	0	0	0.1
MAKELA4	21	40	PSG	2	3	3	0	0.3

when applying the preconditioner (HS105, HS111 y HS112). In some problems using the preconditioner reduces iteration count but increments overall cost (EXTRASIM, HS41 y HS63). This last aspect indicates that fine-tuning is required in order to avoid premature activation of the preconditioner. Premature activation of  $P$  does not increase total iteration count but does have a negative impact of overall computational cost. The idea is to find a compromise between updating  $P$  too frequently and iterating a large amount of times. A conservative strategy would be to delay the activation of the preconditioner up until the iterates are *close* to a solution. Another alternative could be to activate preconditioning as soon as the iterates start to be *very feasible* thus promoting the acceptance of the preconditioned direction.

## 5. Concluding remarks

An acceleration scheme for the Augmented Lagrangian method was presented. The associated preconditioner ( $P$ ) exploits the explicit form of the Augmented Lagrangian Hessian ( $H$ ) without estimating its inverse. The strategy is modular and uses two main ingredients. An auxiliary preconditioner ( $P_M$ ) associated with the Lagrangian Hessian ( $M$ ) and a storage matrix ( $B$ ) related to the constraints Jacobian matrix and possible rank-one corrections of  $H$ . The preconditioner takes inspiration in the Sherman-Morrison identity and Miller's inverse formula. A virtue of this scheme is that the acceleration strategy is agnostic to the class of preconditioner used for  $P_M$ . For special choices of approximations to the Lagrangian Hessian, the preconditioner is the exact inverse. Quasi-Newton and Secant-type approximations are encouraged, associated low-rank BFGS corrections are absorbed in the  $B$  matrix and do not require explicit handling. The quality of  $P$  is determined by the quality of  $P_M$  and the relaxation over the constraints during the assembly of  $B$ . Some characteristics of  $P$  are induced by  $P_M$ . The explicit handling of the constraints is attractive when  $m < n$ . This is not necessarily the case when  $m \approx n$  or  $m > n$ , since  $VV^T$  is possibly rank complete and the paradigm of using the explicit form of all constraints is no longer practical. Other methods that take advantage of the aggregated form of the Gauss-Newton matrix are recommended. From the implementation point of view, the preconditioner along with the updating strategies can be directly used on Newton-type as well as on Projected Gradient methods. This reduces the coding effort and allows code recycling.

Numerical experiments gave insight on the quality of the preconditioner and general behavior of the scheme inside ALM. Initial results reveal that when  $P_M$  is poor, so is  $P$ . Optimal assembly parameters are problem specific and not universal. Constraint relaxation is quite delicate and required careful handling. Arbitrary relaxation proved

Table 7: Numerical results for solving convex constrained problems using ALM coupled with SPG and PSPG. Column labels are described in the caption of Table 5.

Name	$n$	$m$	Method	ItL.	Itin	AcM	AcV	Time (s)
AIRPORT	84	42	SPG	n/c	—	—	—	—
AIRPORT	84	42	PSPG	81	304	159	0	18.9
AIRPORT	84	42	PSPG	81	304	159	0	22.9
EXTRASIM	2	1	SPG	1	16	—	—	0.2
EXTRASIM	2	1	PSPG	1	2	2	0	0.2
HS41	4	1	SPG	4	78	—	—	0.4
HS41	4	1	PSPG	4	78	6	0	0.7
HS63	3	2	SPG	4	125	—	—	0.5
HS63	3	2	PSPG	4	26	16	0	0.5
HS90	4	1	SPG	1	10	—	—	0.1
HS90	4	1	PSPG	1	8	1	0	0.2
HS105	8	1	SPG	2	505	—	—	4.5
HS105	8	1	PSPG	1	100	82	0	2.1
HS105	8	1	PSPG	1	80	49	0	1.7
HS111	10	3	SPG	12	5999	—	—	21.7
HS111	10	3	PSPG	12	56	41	0	0.8
HS111	10	3	PSPG	12	86	36	0	1.2
HS112	10	3	SPG	12	1567	—	—	5.7
HS112	10	3	PSPG	12	37	35	0	0.8
LOOTSMA	3	2	SPG	1	34	—	—	0.6
LOOTSMA	3	2	PSPG	1	34	1	0	0.5

to be a poor choice. Recommended relaxation strategies take into account infeasibility measure and gradient size. Further study on the ordering of constraints during the assembly of  $B$  is recommended. During the experimentation a wide range of incomplete direct and inverse factorizations were used for the auxiliary preconditioner  $P_M$ . Results show expected behavior and highlight the agnostic quality of the scheme. Good preconditioner update strategies monitor the changes between iterates. Refreshing on a fixed number of iterations is not recommended. The acceleration scheme was successfully coupled to two inner solvers: a Truncated-Newton machinery for unconstrained sub-problems and a projected gradient-type for convex constrained problems. From Tables 5, 6 and 7 it can be observed that inside the ALM context, preconditioning is not only attractive, but can also be the only alternative that produces a solution. Also can be inferred that attractive assembly parameter values, relaxation and update tolerances for Quasi-Newton machinery are also good for the Spectral Projected Gradient case.

The natural next step is the implementation of the scheme in a low level language and incorporation inside an existing ALM implementation with the objective of further experimenting on larger problems.

## References

- [1] E. G. Birgin, J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM, Philadelphia, USA, 2014.
- [2] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, J. M. Martínez, S. A. Santos, Augmented Lagrangian Algorithms Based on the Spectral Projected Gradient Method for Solving Nonlinear Programming Problems, *Journal of Optimization Theory and Applications* 123 (3) (2004) 497–517.
- [3] R. Andreani, E. G. Birgin, J. M. Martínez, M. L. Schuverdt, On augmented Lagrangian methods with general lower-level constraints, *SIAM Journal on Optimization* 18 (4) (2007) 1286–1309.
- [4] R. Andreani, E. G. Birgin, J. M. Martínez, M. L. Schuverdt, Augmented Lagrangian methods under the constant positive linear dependence constraint qualification, *Mathematical Programming* 111 (1–2) (2008) 5–32.
- [5] R. T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, *SIAM Journal on Control* 12 (2) (1974) 268–285.
- [6] N. Krejić, J. M. Martínez, M. Mello, E. A. Pilotta, Validation of an Augmented Lagrangian Algorithm with a Gauss-Newton Hessian Approximation Using a Set of Hard-Spheres Problems, *Computational Optimization and Applications* 16 (3) (2000) 247–263.
- [7] A. R. Conn, N. I. M. Gould, P. L. Toint, A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM Journal on Numerical Analysis* 28 (2) (1991) 545–572.
- [8] J. C. Simo, T. A. Laursen, An augmented Lagrangian treatment of contact problems involving friction, *Computers & Structures* 42 (1) (1992) 97–116.
- [9] M. Fortin, R. Glowinski, *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems*, Elsevier, Amsterdam. Holanda, 2000.
- [10] M. V. Afonso, J. M. Bioucas-Dias, M. A. T. Figueiredo, An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems, *IEEE Transactions on Image Processing* 20 (3) (2011) 681–695.
- [11] M. Benzi, M. A. Olshanskii, Z. Wang, Modified augmented Lagrangian preconditioners for the incompressible Navier-Stokes equations, *International Journal for Numerical Methods in Fluids* 66 (4) (2011) 486–508.
- [12] T. Heister, G. Rapin, Efficient augmented Lagrangian-type preconditioning for the Oseen problem using Grad-Div stabilization, *International Journal for Numerical Methods in Fluids* 71 (1) (2013) 118–134.
- [13] M. Benzi, M. A. Olshanskii, Field-of-values convergence analysis of augmented Lagrangian preconditioners for the linearized Navier-Stokes problem, *SIAM Journal on Numerical Analysis* 49 (2) (2011) 770–788.
- [14] N. I. M. Gould, D. Orban, P. L. Toint, GALAHAD, a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization, *ACM Transactions on Mathematical Software (TOMS)* 29 (4) (2003) 353–372.
- [15] A. R. Conn, N. I. M. Gould, P. L. Toint, *LANCELOT: a Fortran package for large-scale nonlinear optimization (Release A)*, Springer, Berlin. Alemania, 2013.
- [16] E. G. Birgin, J. M. Martínez, Structured minimal-memory inexact quasi-Newton method and secant preconditioners for Augmented Lagrangian Optimization, *Computational Optimization and Applications* 39 (1) (2008) 1–16.
- [17] M. R. Hestenes, Multiplier and gradient methods, *Journal of Optimization Theory and Applications* 4 (5) (1969) 303–320.
- [18] R. T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming, *Journal of Optimization Theory and Applications* 12 (6) (1973) 555–562.
- [19] E. G. Birgin, J. M. Martínez, M. Raydan, Nonmonotone spectral projected gradient methods on convex sets, *SIAM Journal on Optimization* 10 (4) (2000) 1196–1211.
- [20] E. G. Birgin, J. M. Martínez, M. Raydan, Algorithm 813: SPG—software for convex-constrained optimization, *ACM Transactions on Mathematical Software (TOMS)* 27 (3) (2001) 340–349.
- [21] E. G. Birgin, J. M. Martínez, M. Raydan, Spectral projected gradient methods: review and perspectives, *J. Stat. Softw* 60 (3).
- [22] L. Bello, M. Raydan, Preconditioned Spectral Projected Gradient Method on Convex Sets, *Journal of Computational Mathematics* 23 (3) (2005) 225–232.
- [23] E. G. Birgin, J. M. Martínez, L. Martínez, G. B. Rocha, Sparse projected-gradient method as a linear-scaling low-memory alternative to diagonalization in self-consistent field electronic structure calculations, *Journal of Chemical Theory and Computation* 9 (2) (2013) 1043–1051.
- [24] J. Nocedal, S. Wright, *Numerical optimization*, Springer, Berlin. Alemania, 2006.
- [25] S. Bellavia, V. De Simone, D. di Serafino, B. Morini, Updating Constraint Preconditioners for KKT Systems in Quadratic Programming Via Low-Rank Corrections, *SIAM Journal on Optimization* 25 (3) (2015) 1787–1808.
- [26] J. M. Martínez, An extension of the theory of secant preconditioners, *Journal of Computational and Applied Mathematics* 60 (1995) 115–125.
- [27] J. M. Martínez, A theory of secant preconditioners, *Mathematics of Computation* 60 (202) (1993) 681–698.
- [28] A. R. Conn, N. I. M. Gould, P. L. Toint, Convergence of quasi-Newton matrices generated by the symmetric rank one update, *Mathematical Programming* 50 (1–3) (1991) 177–195.
- [29] J. Nocedal, Updating quasi-Newton matrices with limited storage, *Mathematics of Computation* 35 (151) (1980) 773–782.

- [30] N. Andrei, Scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, *Optimization Methods and Software* 22 (4) (2007) 561–571.
- [31] D. Cores, R. Escalante, M. González-Lima, O. Jimenez, On the use of the spectral projected gradient method for support vector machines, *Computational & Applied Mathematics* 28 (2009) 327–364.
- [32] Y. Dai, R. Fletcher, New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds, *Mathematical Programming* 106 (3) (2006) 403–421.
- [33] J. Sherman, W. J. Morrison, Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix, *Ann. Math. Statist.* 21 (1) (1950) 124–127.
- [34] A. M. Sajo–Castelli, Precondicionamiento del Método Lagrangiano Aumentado, Ph.D. thesis, Universidad Simón Bolívar, to be published (2017).
- [35] M. A. Woodbury, Inverting modified matrices, Tech. rep., Statistical Research Group, Princeton University (1950).
- [36] W. W. Hager, Updating the Inverse of a Matrix, *SIAM Review* 31 (2) (1989) 221–239.
- [37] K. S. Miller, On the inverse of the sum of matrices, *Mathematics Magazine* 54 (2) (1981) 67–72.
- [38] M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method, *IMA Journal of Numerical Analysis* 13 (3) (1993) 321–326.
- [39] M. Raydan, The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem, *SIAM Journal on Optimization* 7 (1) (1997) 26–33.
- [40] R. Fletcher, Low storage methods for unconstrained optimization, *Lectures in Applied Mathematics (AMS)* 26 (1990) 165–179.
- [41] R. A. Horn, C. A. Johnson, *Matrix Analysis*, Cambridge University Press, New York City. USA, 1985.
- [42] L. N. Trefethen, D. Bau III, *Numerical Linear Algebra*, SIAM, Philadelphia. USA, 1997.
- [43] S. Babaie-Kafaki, A modified BFGS algorithm based on a hybrid secant equation, *Science China Mathematics* 54 (9) (2011) 2019–2036.
- [44] W. J. Leong, M. A. Hassan, M. Farid, A monotone gradient method via weak secant equation for unconstrained optimization, *Taiwanese Journal of Mathematics* 14 (2) (2010) 413–423.
- [45] J. M. Martínez, A family of quasi-Newton methods for nonlinear equations with direct secant updates of matrix factorizations, *SIAM Journal on Numerical Analysis* 27 (4) (1990) 1034–1049.
- [46] D. Li, M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, *Journal of Computational and Applied Mathematics* 129 (1–2) (2001) 15–35.
- [47] N. I. M. Gould, D. Orban, P. L. Toint, CUTEst: A Constrained and Unconstrained Testing Environment with Safe Threads for Mathematical Optimization, *Comput. Optim. Appl.* 60 (3) (2015) 545–557.
- [48] M. Benzi, C. D. Meyer, M. Tũma, A sparse approximate inverse preconditioner for the conjugate gradient method, *SIAM Journal on Scientific Computing* 17 (5) (1996) 1135–1149.
- [49] M. Benzi, J. K. Cullum, M. Tũma, Robust approximate inverse preconditioning for the conjugate gradient method, *SIAM Journal on Scientific Computing* 22 (4) (2000) 1318–1332.
- [50] M. Benzi, M. Tũma, A robust incomplete factorization preconditioner for positive definite matrices, *Numerical Linear Algebra with Applications* 10 (5–6) (2003) 385–400.
- [51] M. Benzi, D. Bertaccini, Approximate inverse preconditioning for shifted linear systems, *BIT Numerical Mathematics* 43 (2) (2003) 231–244.
- [52] R. F. Boisvert, R. Pozo, K. Remington, R. F. Barrett, J. J. Dongarra, Matrix Market: A Web Resource for Test Matrix Collections, in: *The Quality of Numerical Software: Assessment and Enhancement*, Chapman & Hall, Londres. UK, 1997, pp. 125–137.